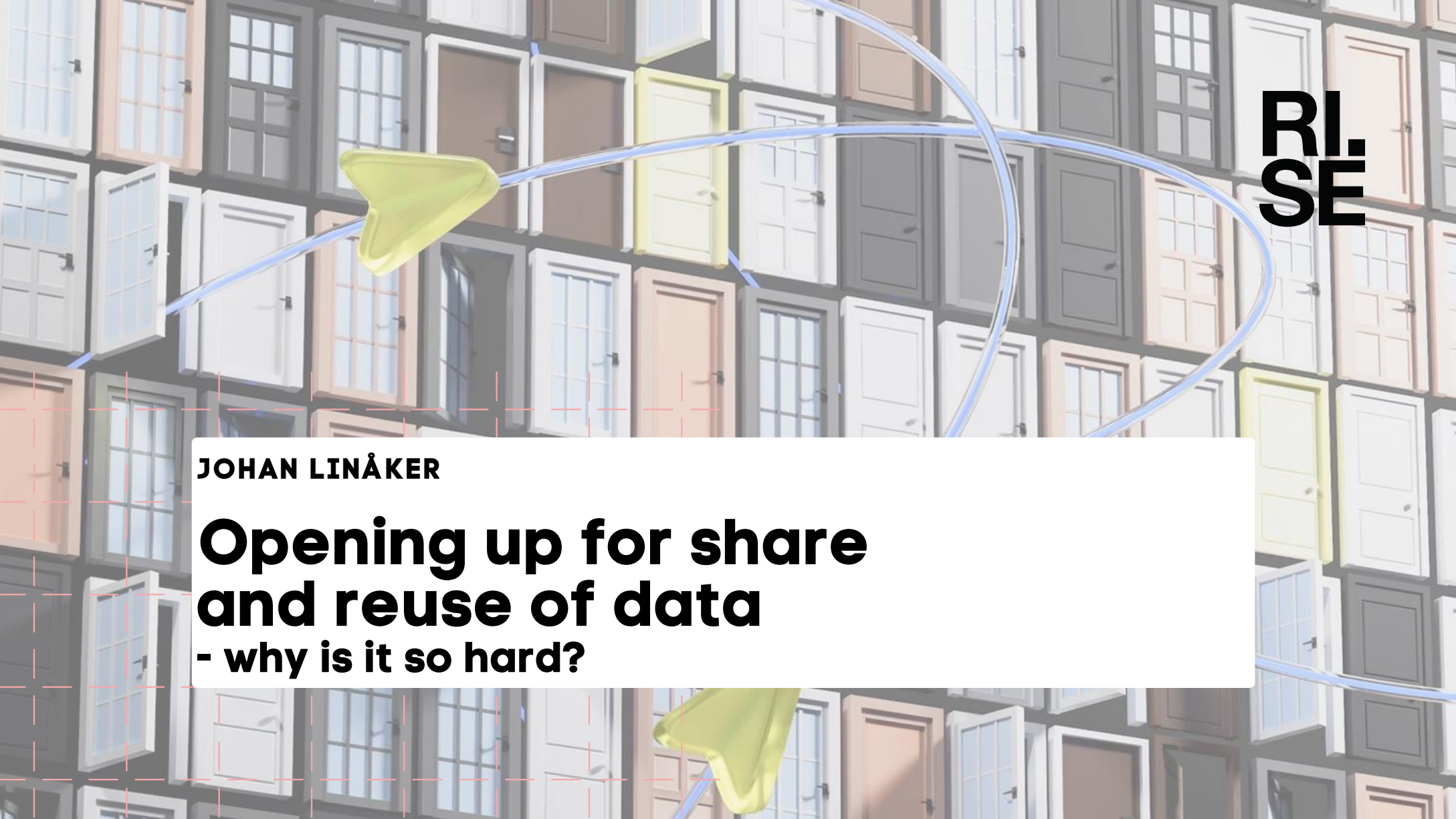


JOHAN LINÅKER

Opening up for share and reuse of data

- why is it so hard?

The Economist @TheEconomist · 2h

The world's most valuable resource is no longer oil, but data

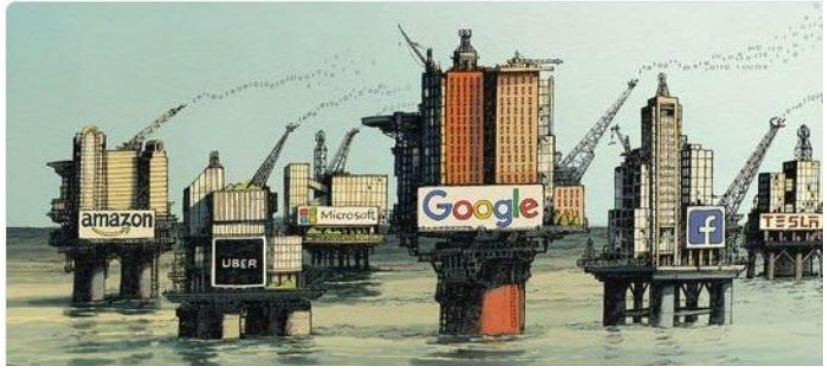


Photo by Bernardo Ferrari | https://unsplash.com/photos/a-large-structure-in-the-middle-of-a-body-of-water-5ZPz4q_X38U

Software the refinery

Photo by tasukaran | https://pixabay.com/photos/factory-night-view-refinery-sunset-7806641/

AI the engine

Photo by k_jpg | https://pixabay.com/photos/motor-cars-speed-automotive-2828878/

But, where is the data?

Getty Images For Unsplash+| https://unsplash.com/photos/close-up-view-of-the-cars-hood-focused-view-of-parts-l-w5DRlnSPA

Need to Open up

Photo by NEOM | https://unsplash.com/photos/a-man-walking-down-a-dirt-road-next-to-a-mountain-IPtSV340-j4

What is Open Data?

- Open data and content that can freely be used, modified and shared, by whomever, for what ever reason*
- Permissive and copyleft licenses similar to that of Open Source Software
- E.g., CC0 → CC-BY → CC-BY-SA
- CC0 recommended for open data by DIGG and PRV

*https://opendefinition.org/

Photo by Anna Hunko | https://unsplash.com/photos/ajE5goOGzZc

Not black or white

- Several quality aspects and nuances impact the openness of the data, including
  - Machine readability
  - Platform independence
  - Accessibility
  - Type of format
  - Cost
  - Level of permissiveness
  - Completeness
  - Orginality
  - Linked

* https://opendefinition.org/
https://opengovdata.org/
https://5stardata.info/en/

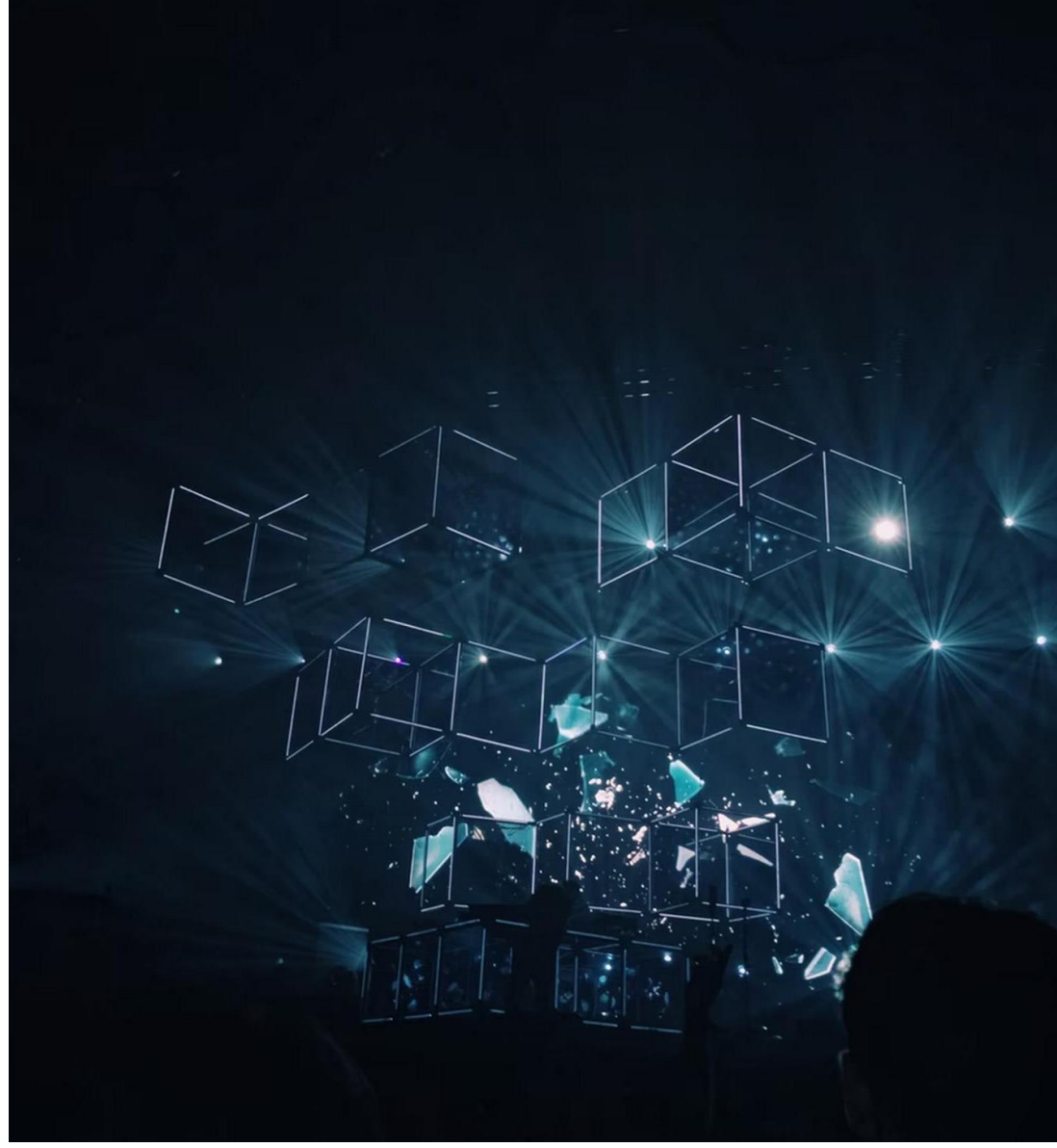


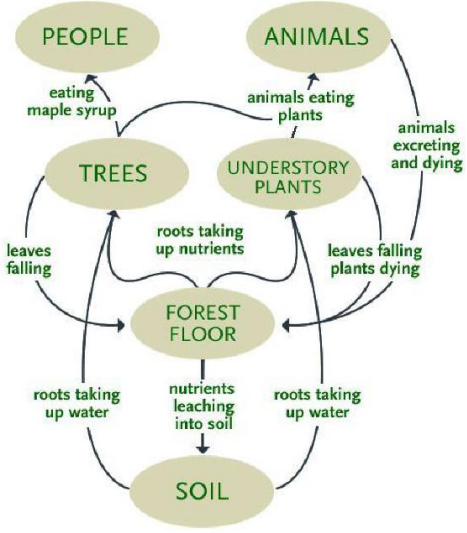
Photo by Fabio | https://unsplash.com/photos/oyXis2kALVg

Collaboration challenges

- Distributed and enriched in value chains from produces to consumer
- Often single lane without feedback loops
- Top-down focus on publishing
- Need for dialogue and interaction
- Should ideally be built as a value network with double pointing edges

Photo by Mitchell Luo | https://unsplash.com/photos/H3htK85wwnU

Ecosystem as a metaphor



@johanlinaker

Photo by Sandy Millar | https://unsplash.com/photos/5ZVL9gVLWv4

Data ecosystems

- An open network of actors, gathering round a joint interest and platform promoting that interest
- Enables use, sharing, and collaboration round the data and related resources in the ecosystem
- Done through the exchange of knowledge, artifacts, and resources

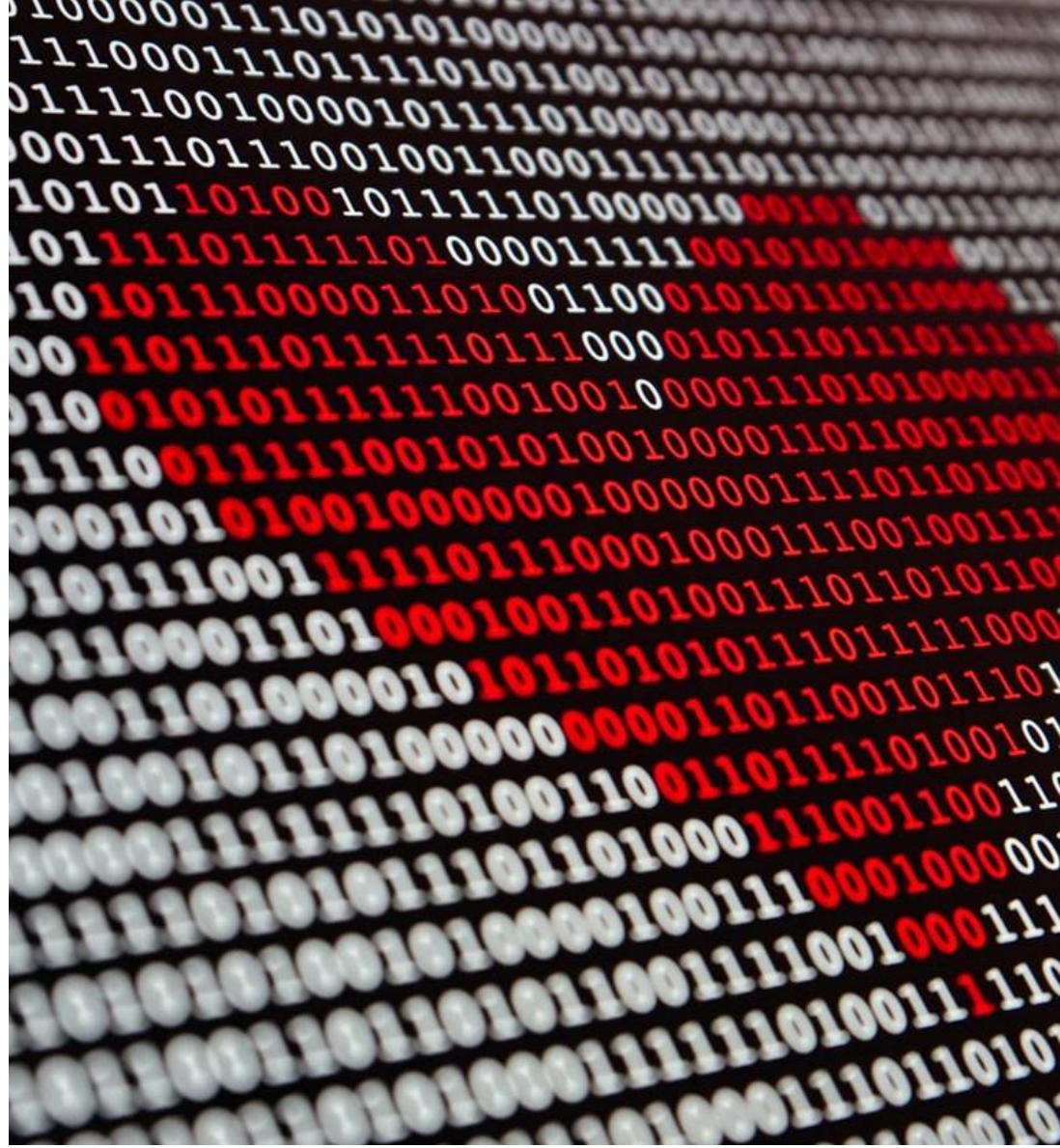


Photo by Denny Müller | https://unsplash.com/photos/HfWA-Axq6Ek

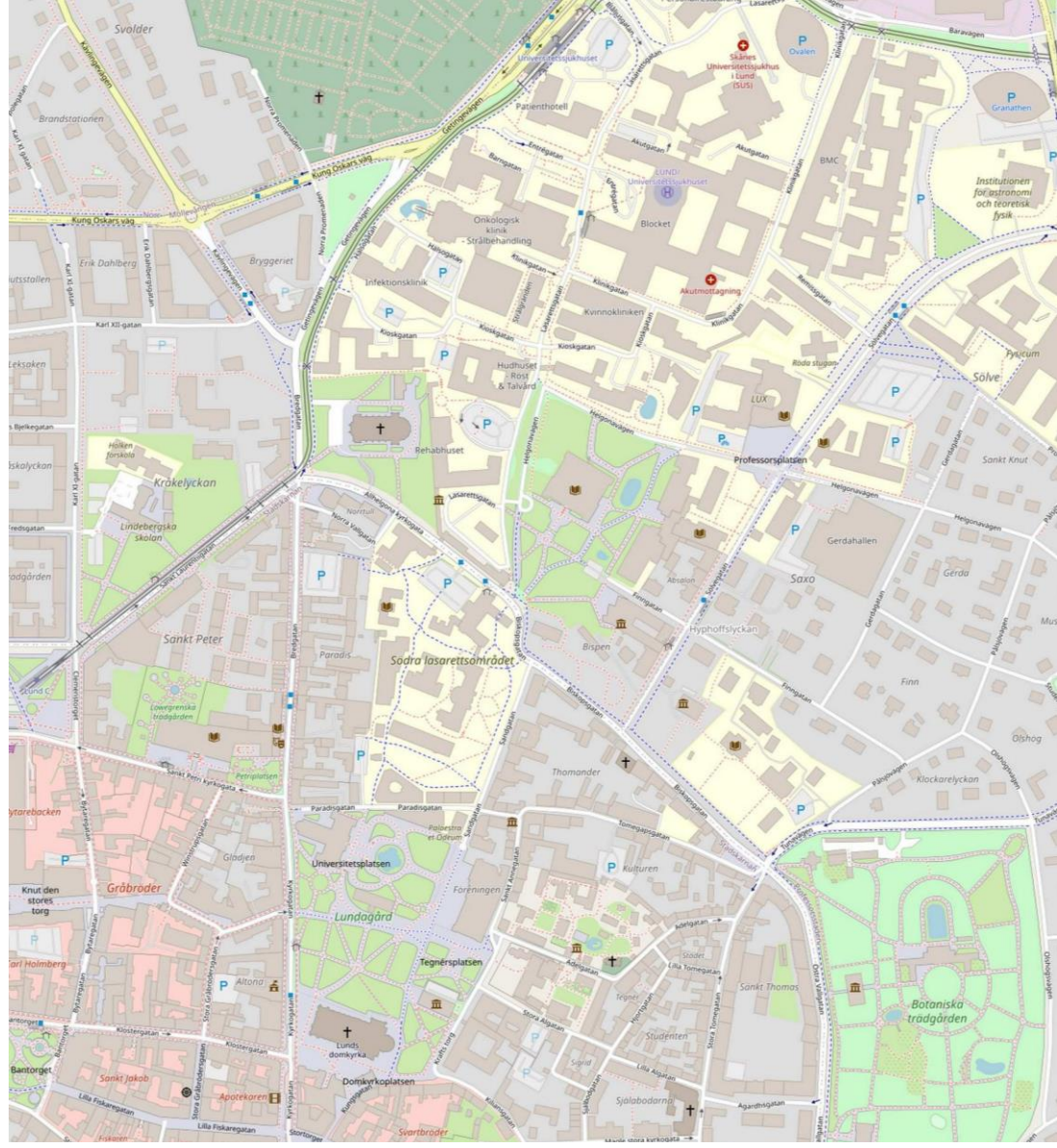
Functional roles

- Data producer
- Data intermediary
- Service provides
- Application developer
- Tools- and infrastructure provides
- ...

Photo by Pascal Swier | https://unsplash.com/photos/7de474KZIbs

Example: OpenStreetMap

- Open map data covering the world with > 1 B contributions
- ”Community of communitys” involving public sector, industry, civil society, academia, hobbyists
- Driven by consencus norms without an explicit hierarchy or orchestrator
- Hosted by a dedicated foundation in the UK
- Open collaboration on the editing of the data and related Open Source Software tools



@johanlinaker | https://linaker.se

Example: Trafiklab

- Orchestrated by Samtrafiken, public entity co-owned by Public Transport Authorities and Private transport operators
- Publishes open data related to the public transport in Sweden, e.g., time tables, bus stops, and disturbances
- Tools, frameworks, and application examples available as Open Source Software

Photo by Artur Tumasjan | https://unsplash.com/photos/-ebp8HkKtrc

Example: Catena-X

- ”Open” data ecosystem for the Automotive industry
- A common standard for the secure exchange of data and information between actors
- Defines guidelines and provides OSS reference implementations for secure data exchange along the automotive value creation chain.
- Data provider defines who gets to use the data, and under what conditions
- Gathers SMEs, OEMs, Tier 1-2-3s…
- Governed through a dedicated industry foundation, The Catena-X Automotive Network e.V.

Photo by Brock Wegner | https://unsplash.com/photos/black-car-gps-turned-on-in-car-pWGUMQSWBwI

Example: European Mobility Dataspace

- Grounded in the European Strategy for Data, aiming to establish a single market for data, ensuring Europe's competitiveness at a global stage.
- Mobility one of several key sectors
- Aims to provide a common technical and governance framework to enable interoperability and remove barriers to data access and sharing in the mobility and transport sector.
- Builds on several existing data ecosystems
- Explored through 8 local implementation sites across Europe, including Stockholm, and a number of use cases

Photo by Brock Wegner | https://unsplash.com/photos/black-car-gps-turned-on-in-car-pWGUMQSWBwI

Sharing is still limited

Photo by Josh Appel | https://unsplash.com/photos/padlock-on-gray-chain-link-fence-n55P36NmaE4

General sharing challenges

- Confidentiality or business aspects
- Legal barriers and uncertainties
- Costs of opening up and maintaining
- Technical competency
- Lacking incentives and culture
- Unethical and unexpected use case
- Unintended competition
- Knowing what data is needed
- Achieving external reuse and collaboration on data

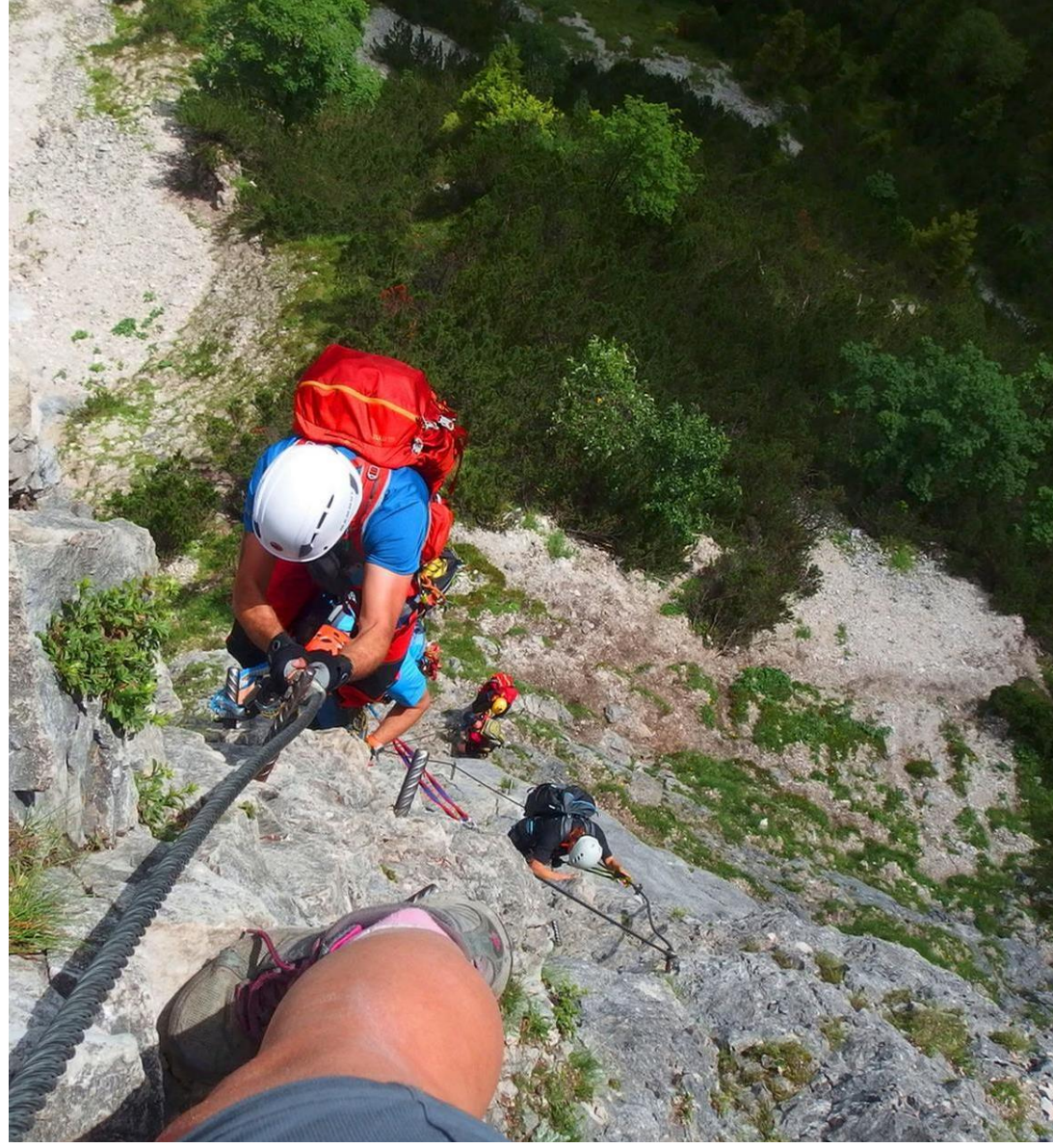


Photo by Maja Kochanowska | https://unsplash.com/photos/E-v6j5Y9xAs

Technical challenges

- Finding and curating high-quality data
- Managing continuous changes in the data through versioning, labelling, and sharing practices
- Data quality assurance and lack of standardized annotation formats
- No or limited availability of adequate tool-support, and infrastructure for hosting data projects

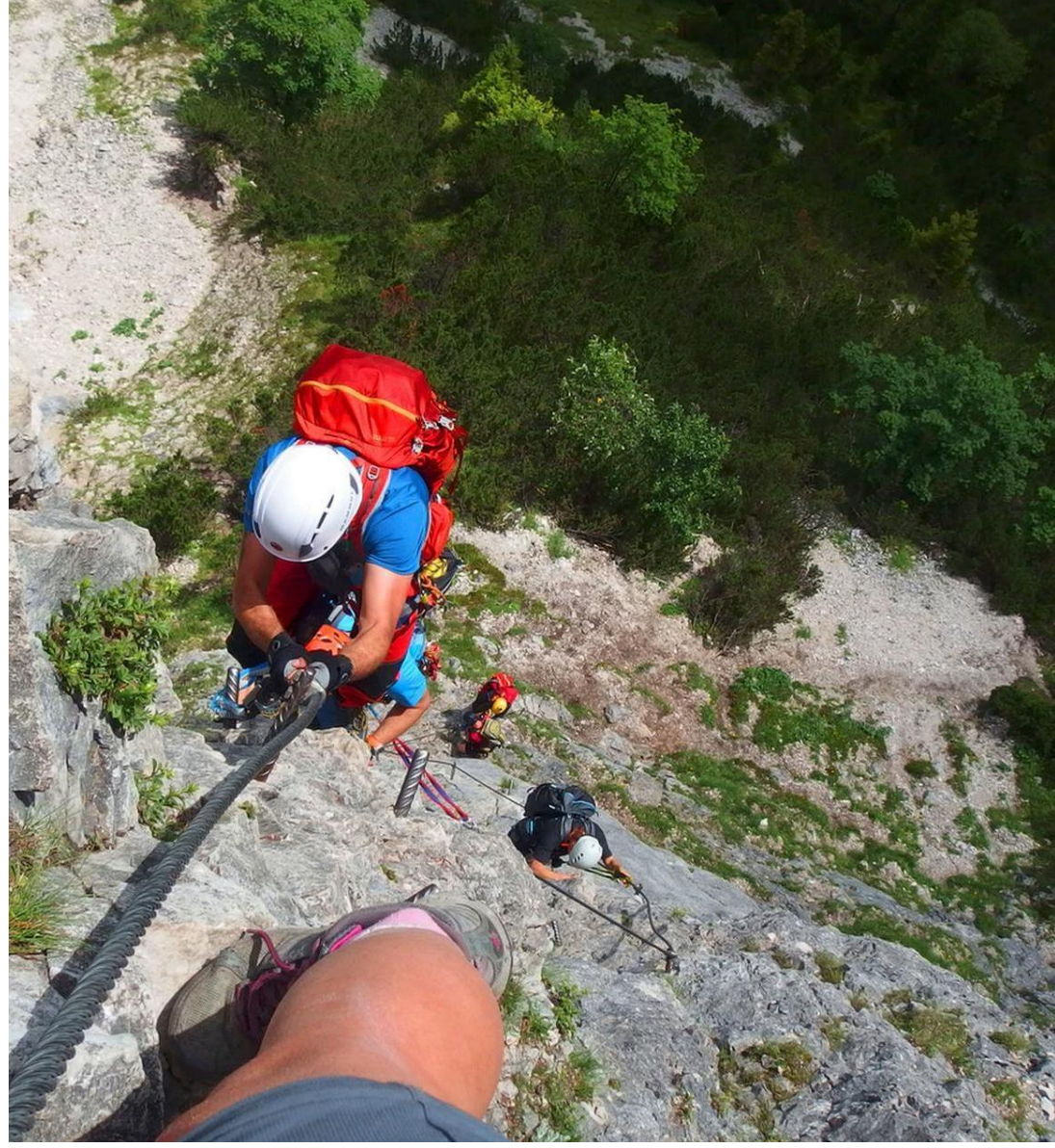


Photo by Maja Kochanowska | https://unsplash.com/photos/E-v6j5Y9xAs

Challenges in the European Mobility Data Space

Dmitrii Eliuseev | https://unsplash.com/photos/a-blue-and-white-train-traveling-over-a-bridge-rT5TLbjV57M

Definition of orchestrating body

- Sustainable orchestrator needed that can manage and facilitate the data sharing and maintain technical platforms, tools, building blocks, standards, procurements, etc. required by the ecosystem actors.
- Extent and openness of governance a balance between inclusiveness and assignment and directives for the orchestrating body.

Photo by Larisa Birta | https://unsplash.com/photos/people-playing-violin-inside-dim-room-slbOcNlWNHA

Control and monitoring

- Sharing of commercial and sensitive data require capabilities for control and monitoring of who uses the data, when, how, etc.
- Data usage control, provider/consumer certification and traceability for all participants requested.
- Important to build trust for data providers that data shared is limited as defined, especially towards, e.g., competitors or third-party service providers.
- Requirements less for Open Data, although still present

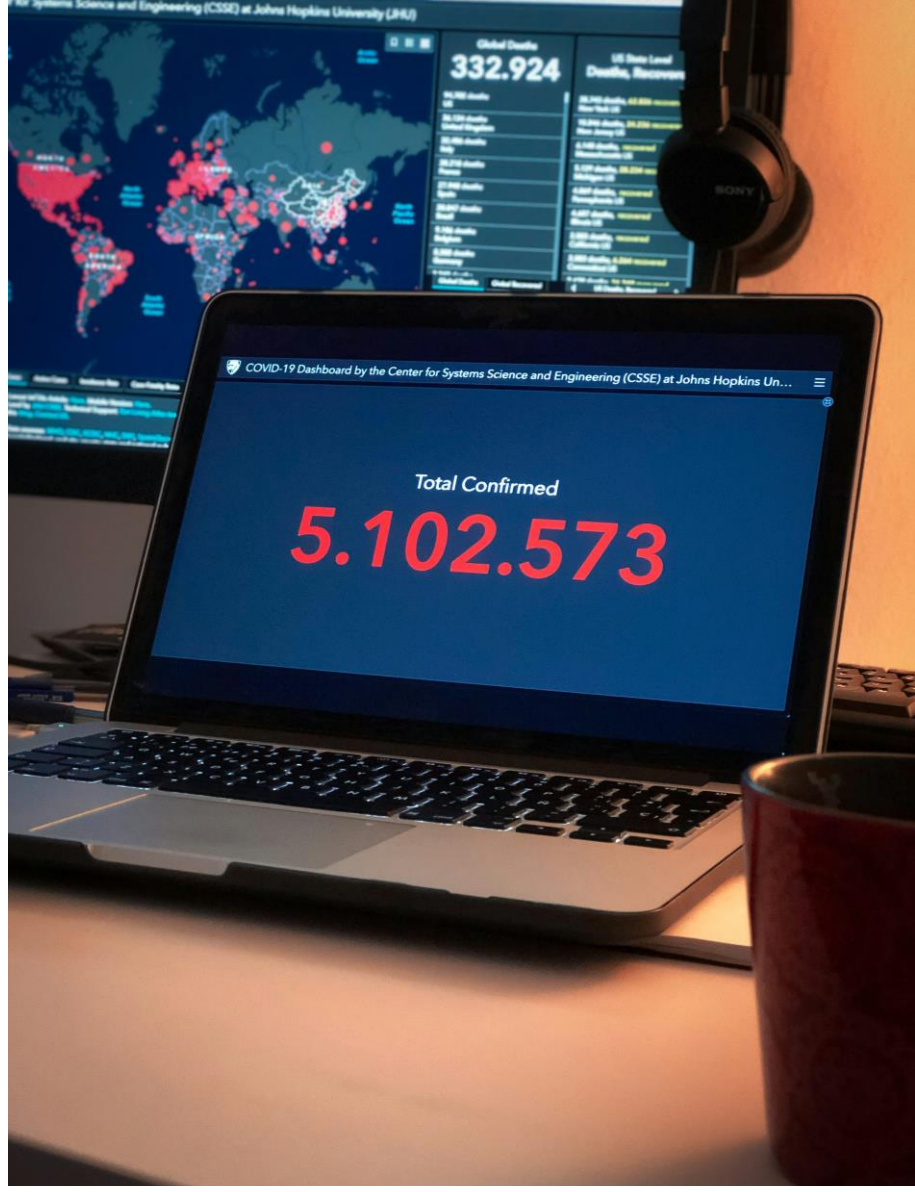


Photo by Isaac Quesada | https://unsplash.com/photos/black-and-silver-laptop-computer-z5k5vWGh7Nk

Business models and services

- Commercial entities need a marketplace where data and payments can be exchanged.
- Business models and cases for third party service or data providers need design, and communication for potential actors.
- Business models need to align with and be enabled by the ecosystem scope and governance.

Photo by Kaleidico | https://unsplash.com/photos/man-wearing-gray-polo-shirt-beside-dry-erase-board-3V8xo5Gbusk

Cross-ecosystem data sharing

- Data providers need to be enabled to share and consume data across borders and ecosystems.
- Coordination and governance required of what data is shared in which ecosystem, and how.
- Third-party service providers need to be enabled to offer their products and services in other markets.

Photo by Markus Spiske | https://unsplash.com/photos/barbed-wire-WWX2bPqP-z4

Scalability in sharing

- Governance and onboarding of new actors need to scale as local data ecosystems open up and connect with other data ecosystems through European data spaces.
- Requirements on control, audit, and traceability of non-open data need to manage actors joining via external ecosystems.

Photo by Markus Spiske | https://unsplash.com/photos/selective-focus-photo-of-plant-spouts-vrbZVyX2k4I

What can we learn from Open Source software?

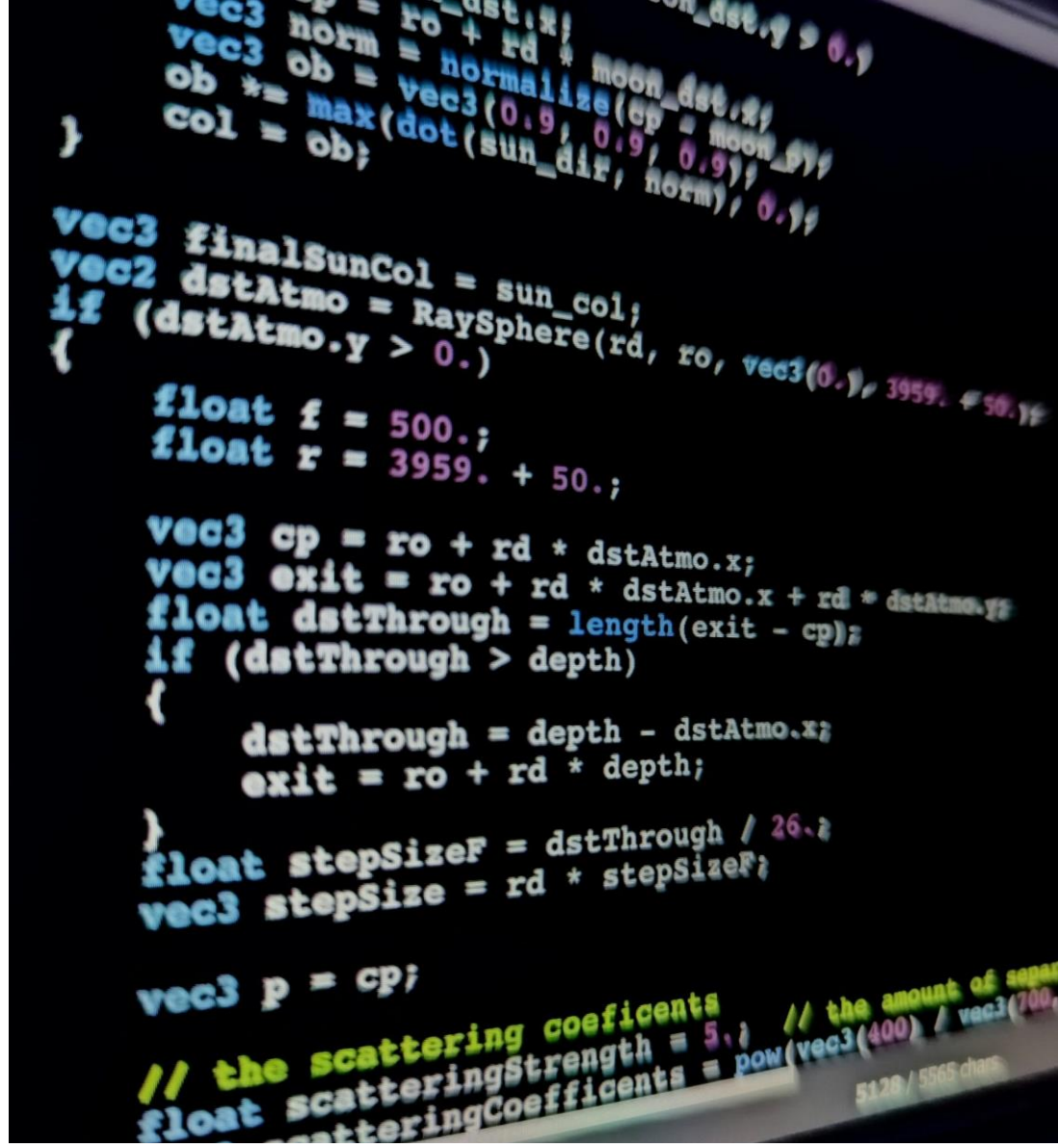


Photo by Andrew | https://unsplash.com/photos/a-screen-shot-of-a-computer-_IBhDwV6PgI

What is Open Source Software?

- Software available under an Open Source Software license
- License that follows the Open Source Definition and is approved by the Open Source Initiative (http://opensource.org)
- Anyone, for whatever reason, may inspect, use, modify the source code and redistribute
- Different conditions apply per license requirements

Photo by Fabian Fauth | https://unsplash.com/photos/1_EedIBc6jY

Open Source Software (OSS) today

- Approximately…
  - 90+ % of all software contains OSS
  - 75% (2020) of companies' code bases consists of OSS (up from 36% 2015)
  - 56 million developers collaborate on OSS projects on GitHub. Estimated to increase > 100 million 2025
  - Collaboration in and between verticals, including Energy, Automotive, Telco, Health

Photo by Pedro Lastra | https://unsplash.com/photos/Nyvq2juw4_o

In other words, it’s everywhere

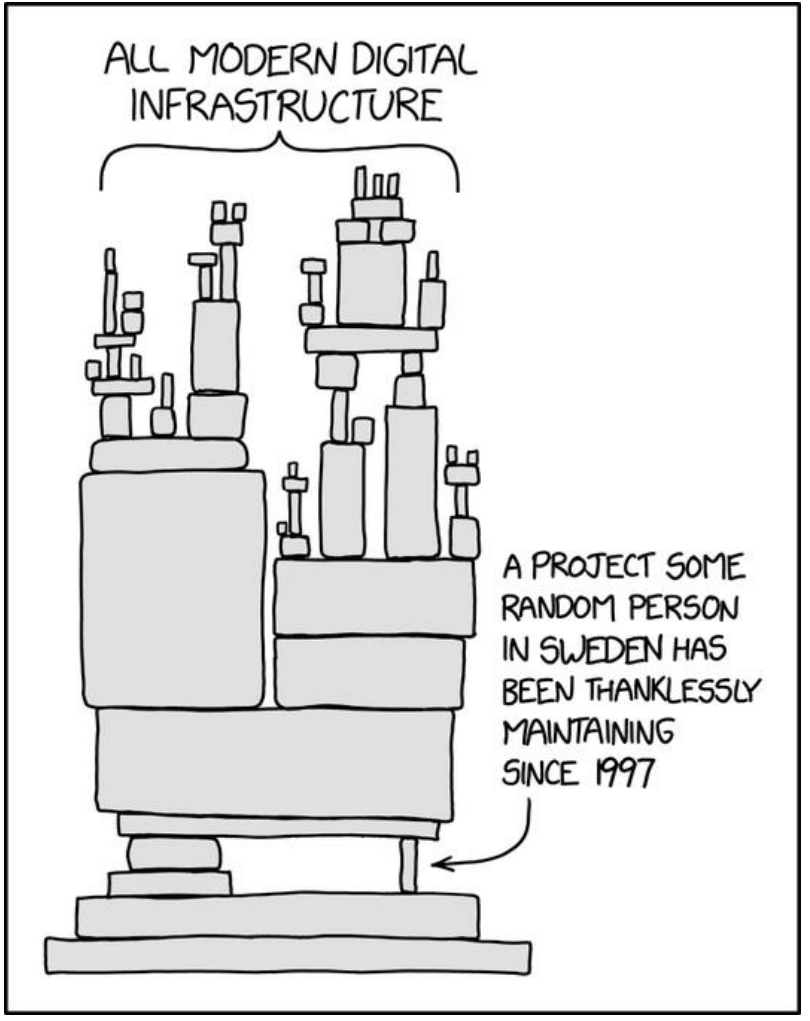


Photo by @tsjost and XKCD | https://twitter.com/tsjost/status/1295662118113619968

Collaboratively developed software

- Software developed as projects by networks of individuals and organizations, aka. Open Source Communities (**or ecosystems**)
- ”Members” of the community commonly both users and developers
- Are united by a common vision and goal around the Open Source Software.

Photo by Annie Spratt | https://unsplash.com/photos/QckxruozjRg

Open development process

- Informal structure pending on community
- Focus is on openness
  - Whoever can contribute
  - Influence through merit
  - Self-appointment of tasks
- Traditional development
  - Carried out in silos
  - Influence though hierarchical status
  - Appointment of tasks

Photo by Leon | https://unsplash.com/photos/Oalh2MojUuk

Open development process

- Transparent and open discussions on bug reports, features, and road map
- Conversations and information persisted in an open infrastructure
- Requirements fragmented and decentralized
- Community full of (un)known stakeholders, all with their own agendas

Photo by Leon | https://unsplash.com/photos/Oalh2MojUuk

Community Structure and Governance

Photo by Leon | https://unsplash.com/photos/Oalh2MojUuk

OSS Project health

- The OSS project’s capability to stay maintained to a high quality, long-term without interruptions
  - Productivity: There is an active development of the project
  - Robustness: The development is open and spread out on several (independent) individuals
  - Openness: Users of the project can influence and contribute to the development of the project

Also comes with its challenges

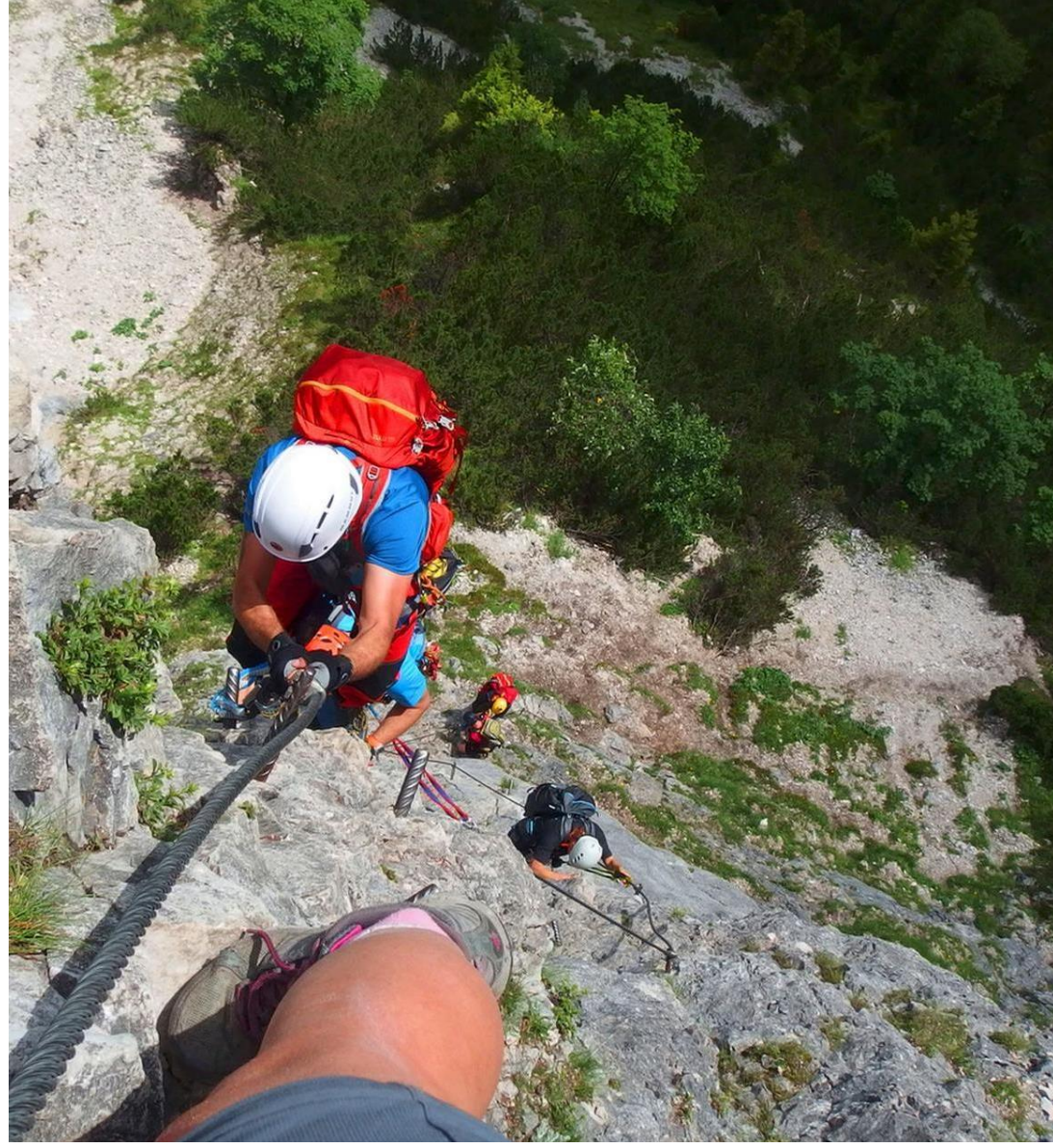
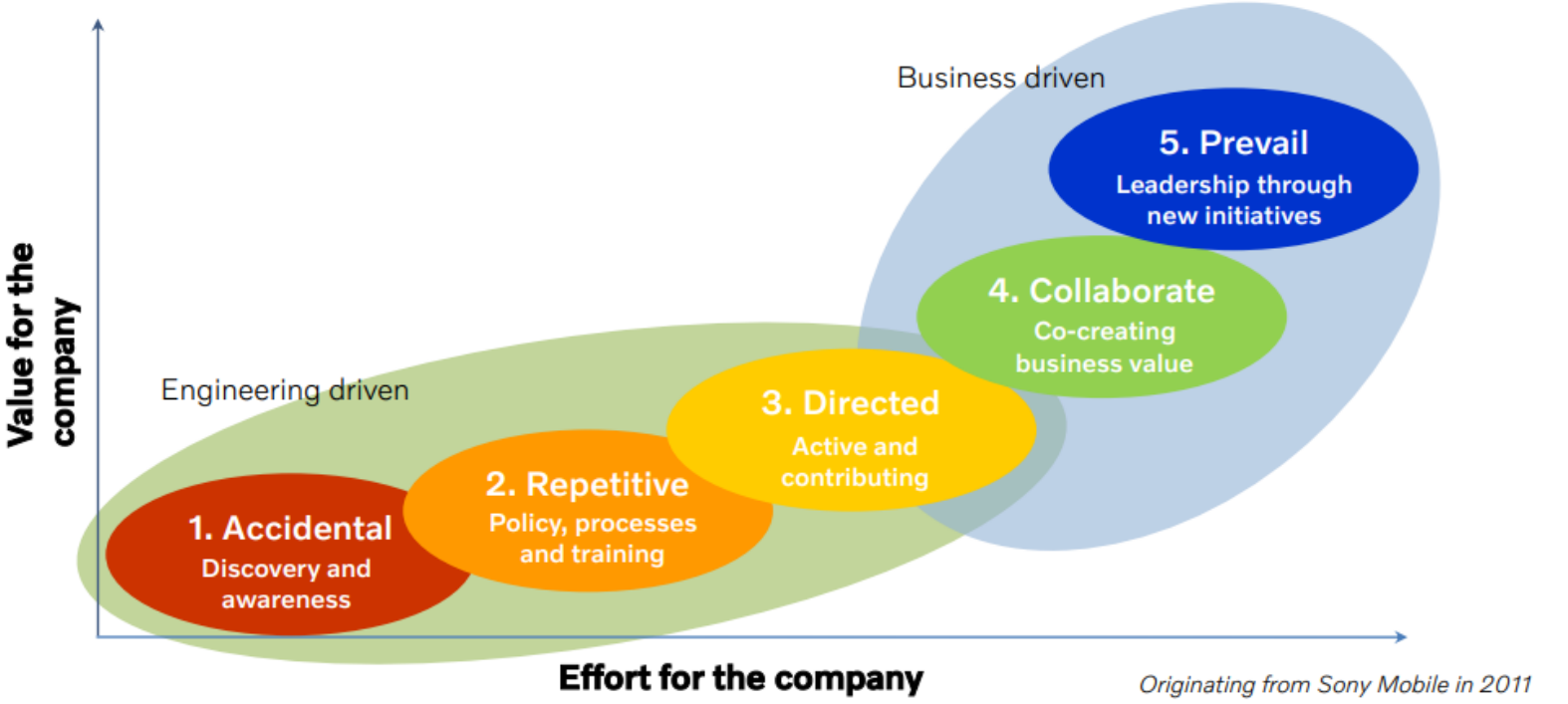


Photo by Maja Kochanowska | https://unsplash.com/photos/E-v6j5Y9xAs

Maturing from consumption to leadership



Originating from Sony Mobile in 2011
/ Adapted by Carl-Eric Mols 2023

RI.SE

Open Source Program Offices (OSPOs)

- Center of competency and support
- Drives organizational readiness and maturity forward on open source
- Designs and executes an organization’s overarching open source strategy
- Provides voice of reason and objectivity on the benefits, risks, and costs of open source and how to balance between
- Supports use, development, and collaboration on open source

Photo by Quino Al | https://unsplash.com/photos/dhGFLj3rI0Q

Deciding what to share

- A balancing act between
  - benefits expected to be gained because of a contribution, and
  - aspects that may complicate the contribution, or in other ways imply cost or risk for the organization

Photo by Katrin Hauf | https://unsplash.com/photos/white-and-brown-weighing-scale-at-0-XyGfxCVbjCA

Incentives for going open source

- Individuals:
  - Sense of belonging,
  - Recognition for contributions,
  - Solves paintpoint,
  - Build CV
- Organizations:
  - Lower costs,
  - Increased innovation,
  - Branding and PR,
  - Strategic tool

Photo by Annie Spratt | https://unsplash.com/photos/QckxruozjRg

Incentives for going open source

- Public policy:
  - Transparency
  - Competition
  - Economic growth
- Researchers:
  - Disseminate research outputs
  - Sustain OSS development between project
  - Collaborate with partners and scientific community
  - Enable reproducibility

Photo by Annie Spratt | https://unsplash.com/photos/QckxruozjRg

Risks, costs and complexities

- Companies:
  - Differentiating functionality, competitive edge and commoditization
  - Sensitive IPR and patents
- Public administrations
  - Compete with industry
  - Ethical aspects and responsibility
  - Integrity and confidentiality
- General:
  - Internal budget and resource constraints
  - Modularity and technical architecture

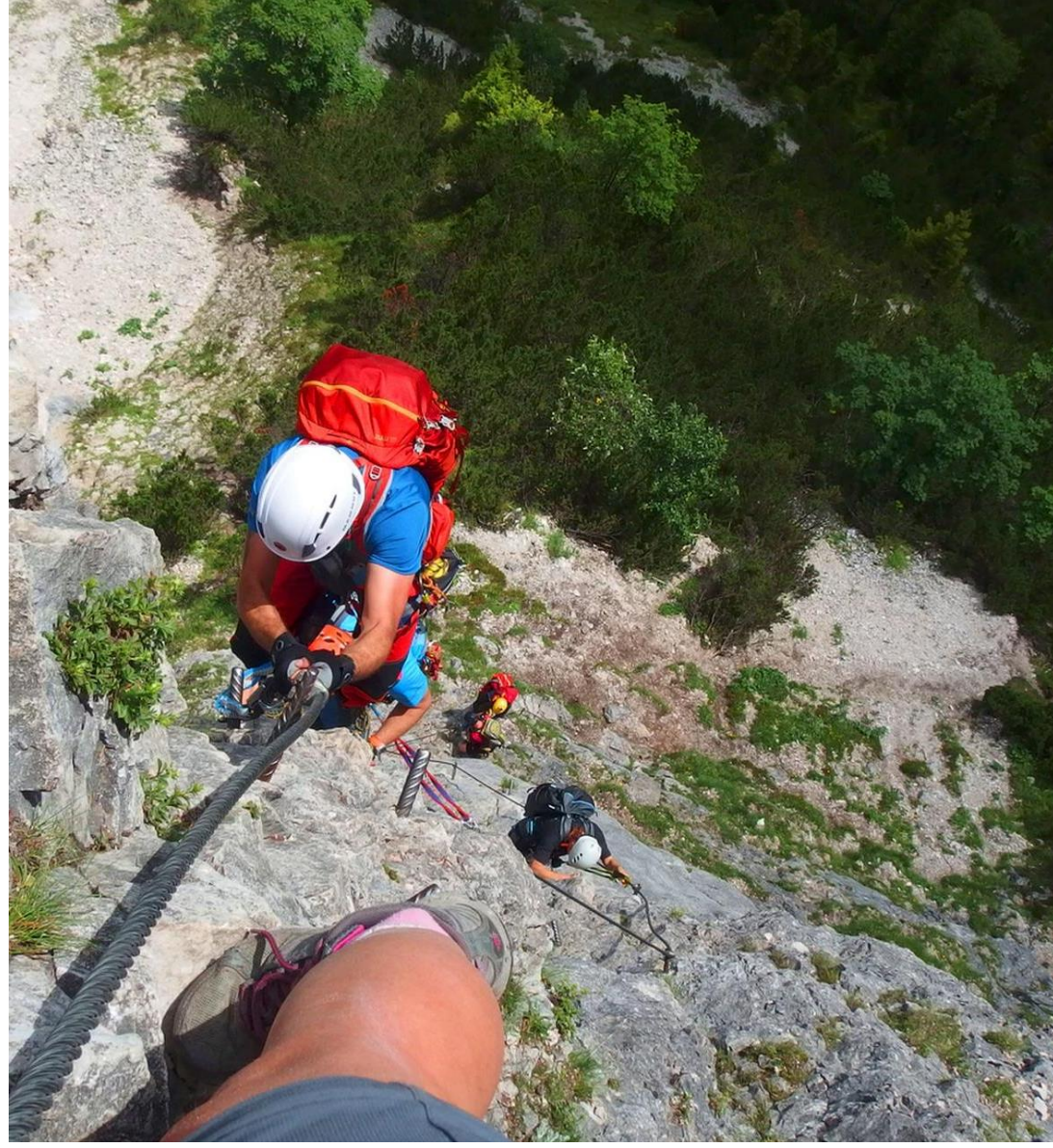
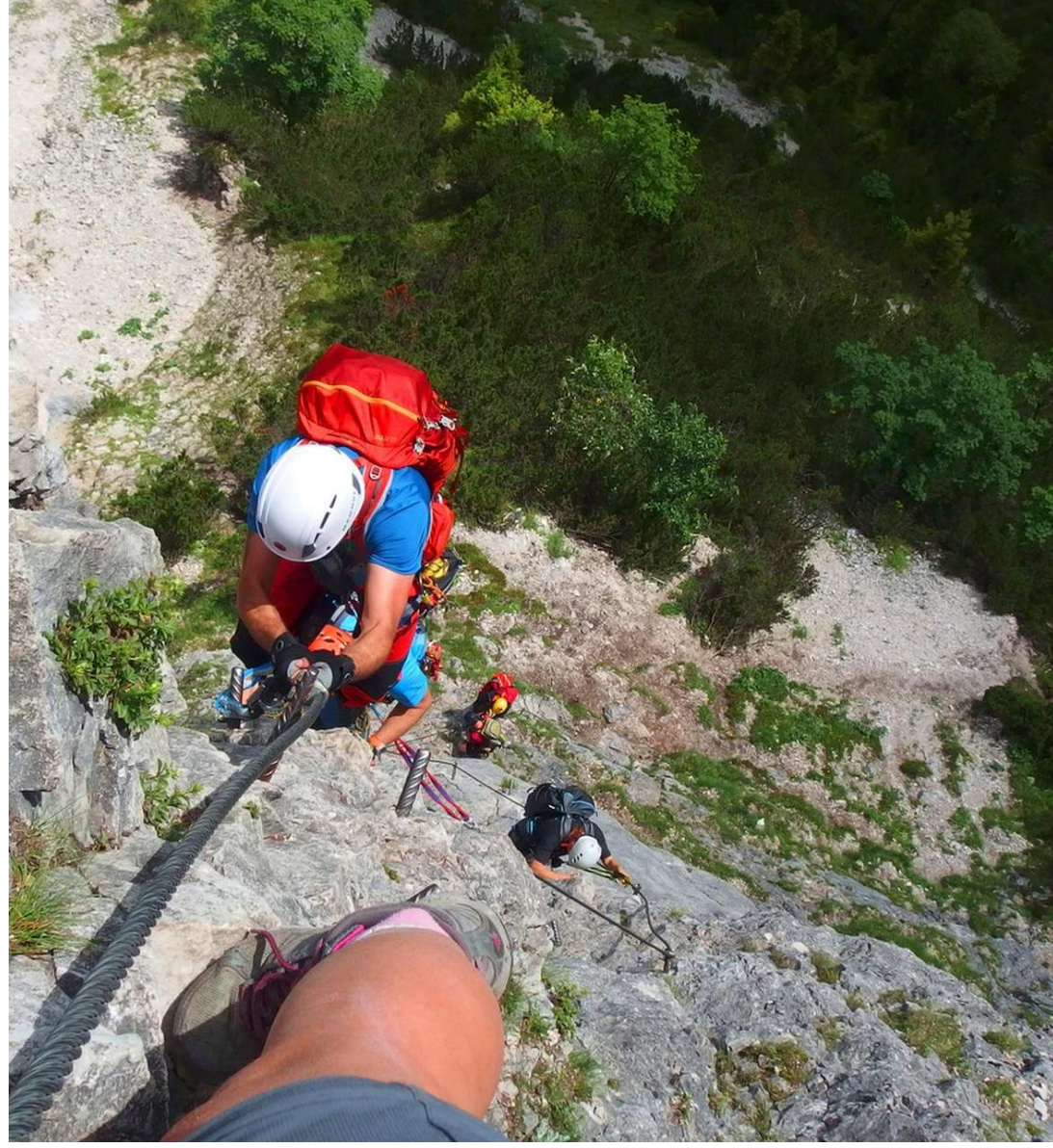


Photo by Maja Kochanowska | https://unsplash.com/photos/E-v6j5Y9xAs

Risks, costs and complexities

- Researchers:
  - Differentiating functionality, competitive edge and commoditization
  - Sensitive IPR and patents
  - Compete with industry
  - Ethical aspects and responsibility
  - Integrity and confidentiality
  - Internal budget and resource constraints
  - Modularity and technical architecture



Open Source AI increases complexity

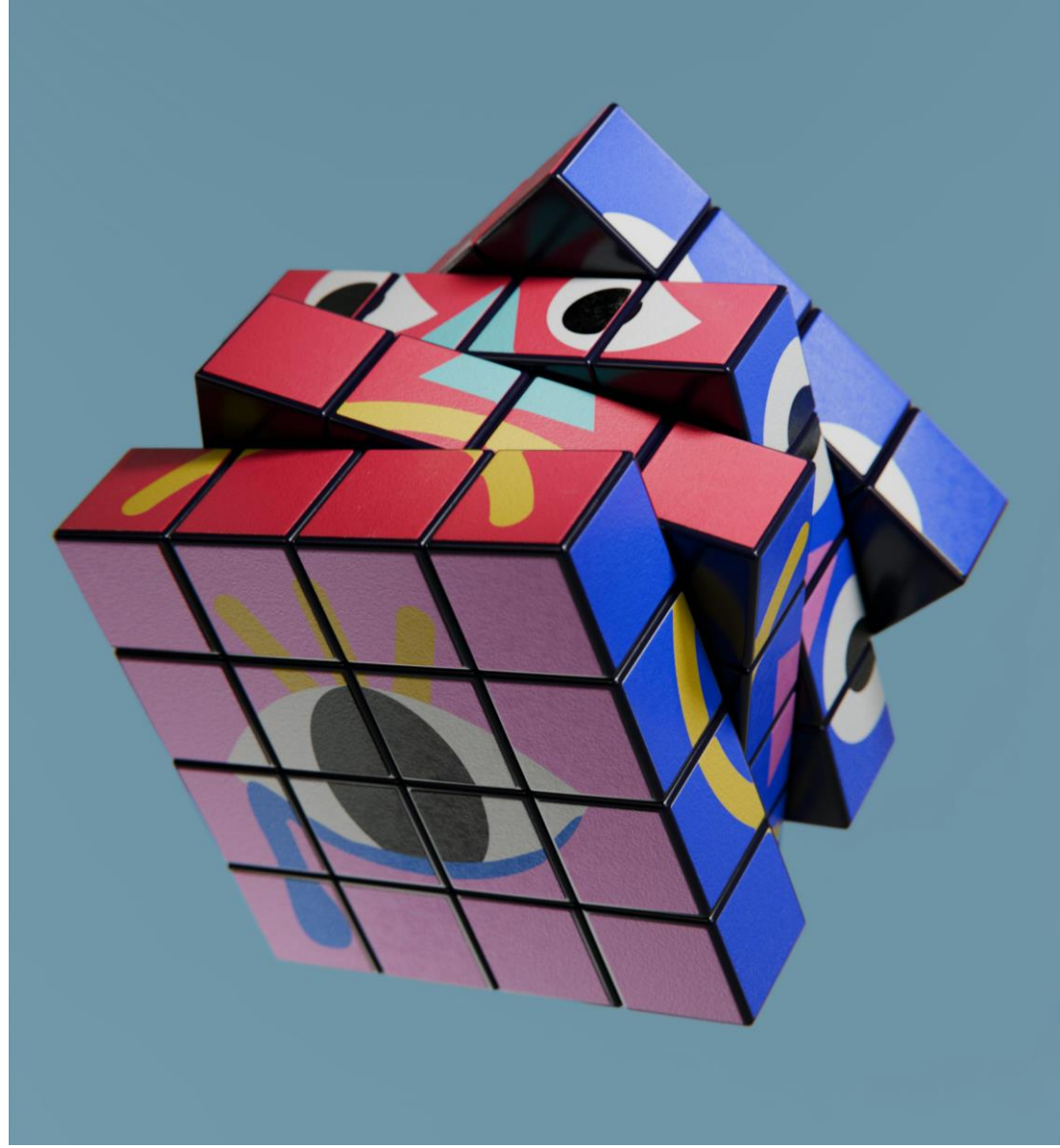


Photo by Luke Jones | https://unsplash.com/photos/a-rubik-cube-flying-in-the-sky-K0znPaPt3tQ

Collaborative development varies

- Presence and form for collaboration may differ based on the component:
  - data (e.g., for training, validation, and testing),
  - source code (e.g., for training and inference),
  - model architecture (e.g., for design choices and hyperparameters), and
  - documentation (e.g., for training procedure and evaluation).

Photo by Desola Lanre-Ologun | https://unsplash.com/photos/woman-and-man-sitting-in-front-of-monitor-IgUR1iX0mqM

Complexity in development

- Many components needed
- Development is costly, e.g.,
  - Collecting and processing data, and
  - Training the model
- Usually limited to resourceful, or venture-backed firms or research institutes

Photo by Scott Blake | https://unsplash.com/photos/seven-construction-workers-standing-on-white-field-x-ghf9LjrVg

Single-vendor vs. community models

- A sliding scale without set definitions
  - Big tech: Llama by Meta,
  - Startups: Mistral, Aleph Alpha
  - Research Institutes: Falcon by Technology Innovation Institute, OLMo by AI2
  - ”Community”: ElutherAI (heavily backed) and BigScience Workshop (Hugging Face)

Photo by Clay Banks | https://unsplash.com/photos/white-and-orange-plastic-chairs-odoykoPOF3E

Open Science and Academia

Photo by Vlad Tchompalov | https://unsplash.com/photos/group-of-people-with-signages-nKNrOZ5MXZY

Common challenges for OSS and data

- Culture, knowledge and organizational support for OSS and open data lacking
  - E.g., license selection, business models, communiy growth…
- Growing sustainable funding for the OSS and data projects' development and maintennace
- Typically very technical and specific knowledge required to contribute and share
- Narrow groups of end-users
- Parallell academic hiearchy inhibiting open collaboration and governnace

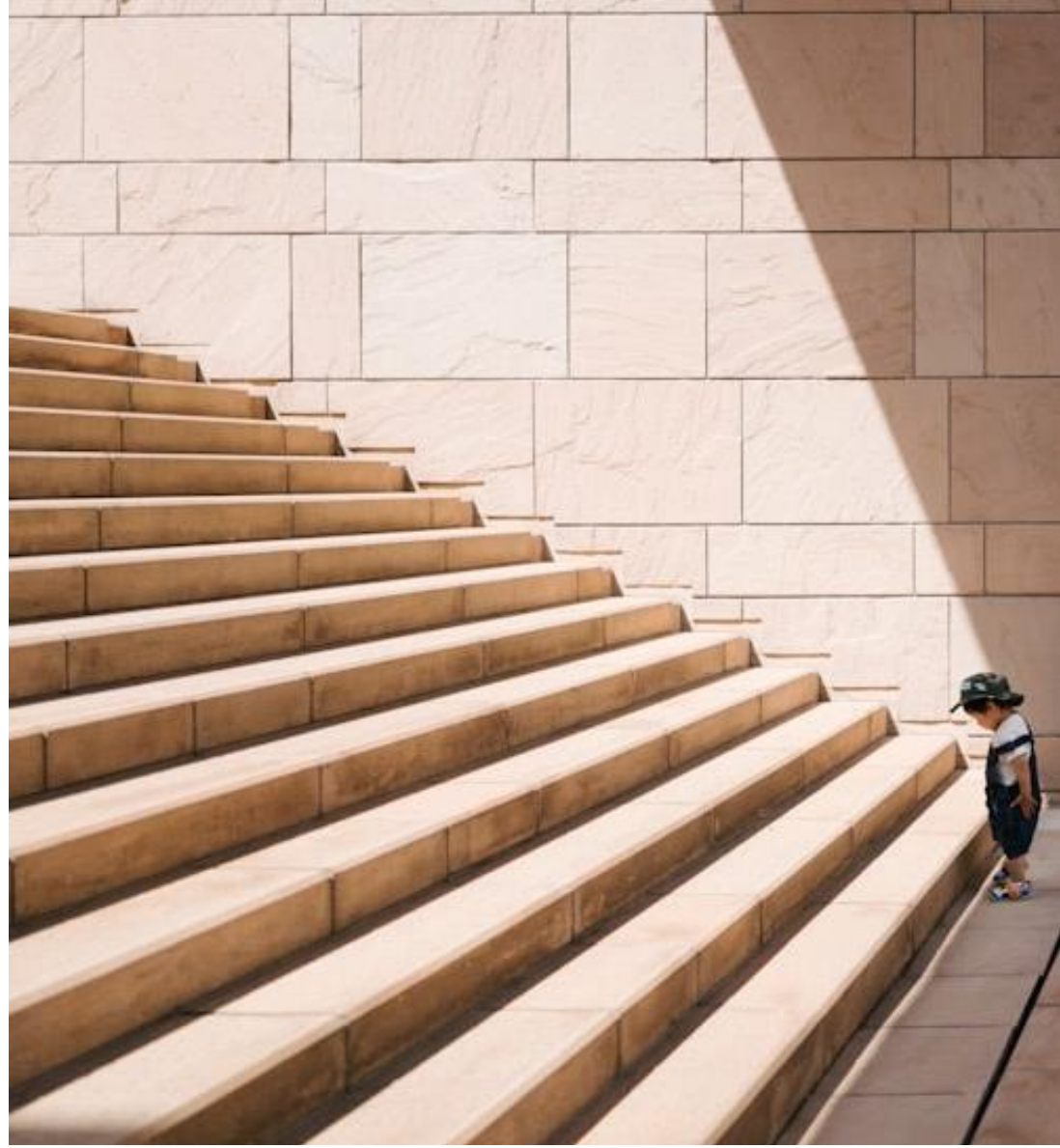


Photo by Jukan Tateisi | https://unsplash.com/photos/toddlers-standing-in-front-of-beige-concrete-stair-bJhT_8nbUA0

Academic OSPOs: Trinity College Dublin

- Small team within the Technology Transfer Office with a business developer and legal expert
- Focused on supporting researchers in using OSS as part of a business model through the commercialization of research outputs
- Supports grant writing and IPR management in research consortiums
- Provides education and training to researchers and under grad students (to various degrees)

Photo by Jonathan Singer | https://unsplash.com/photos/library-interioe-NgU7IJ5XuyY

Academic OSPOs: LERO

- Constituted by an internal community of subject matter experts
- Supports and trains researchers in how to develop, collaborate and disseminate software-based research-outputs as open source
- Considers open source as an instrument for open science, with a broadening interest for other areas within
- Ambition of extending the OSPO and open source as an instrument to the Technology Transfer Office, similar as to Trinity College Dublin

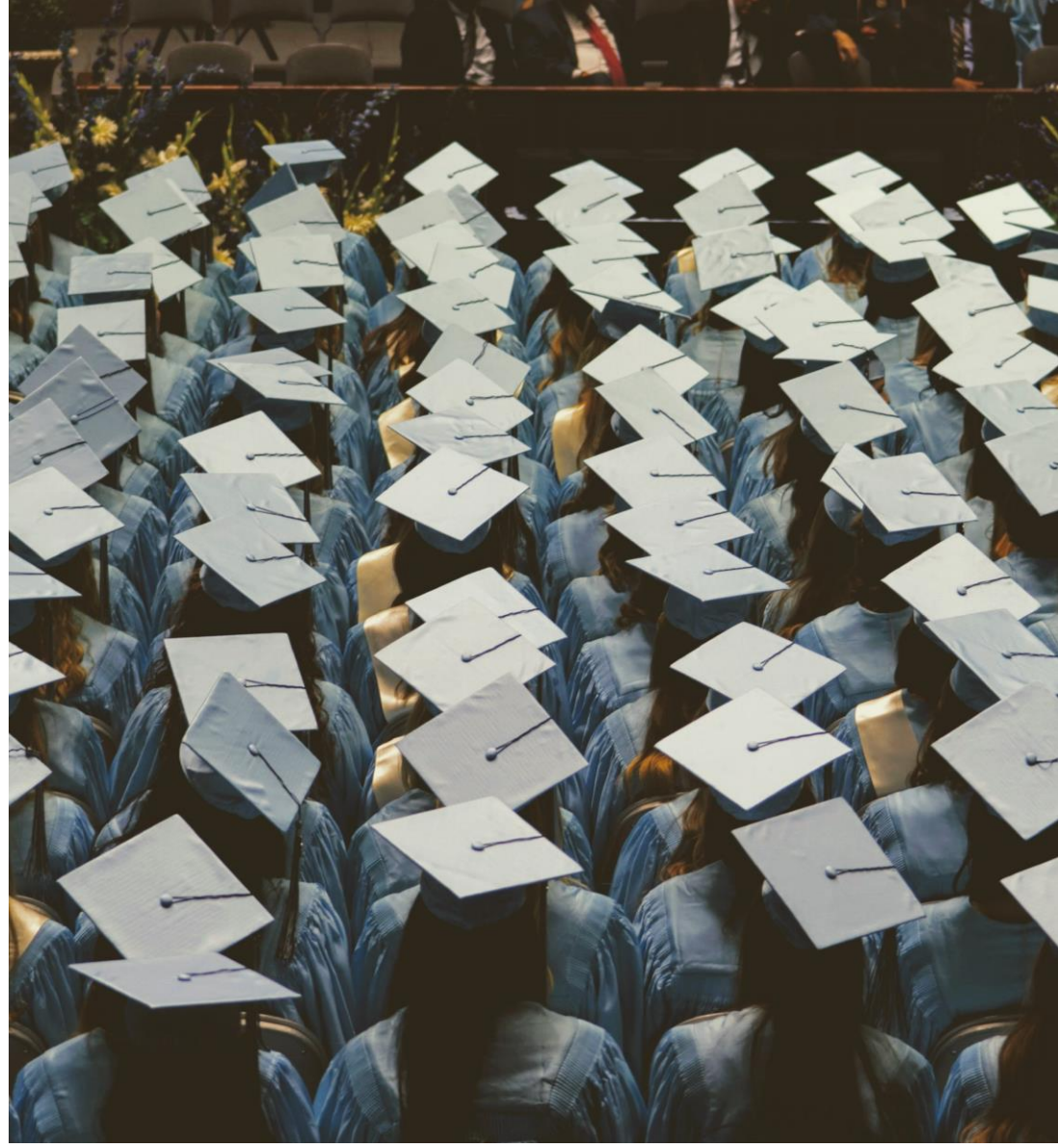
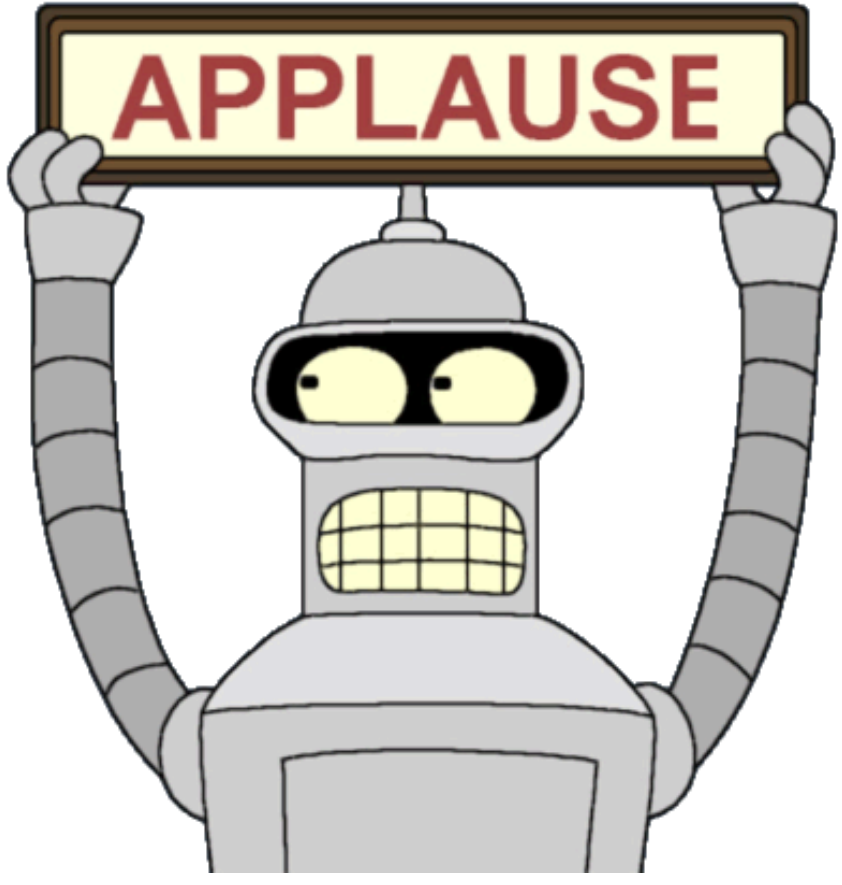


Photo by Joshua Hoehne | https://unsplash.com/fr/photos/cappello-accademico-blu-e-bianco-iggWDxHTAUQ



RI.
SE